Audio Engineering Society

Convention Paper

Presented at the 117th Convention
2004 October 28–31 San Francisco, CA, USA

Spatial Audio Coding: Next-generation efficient and compatible coding of multi-channel audio

J. Herre [1], C. Faller [2], S. Disch[1], C. Ertel[1], J. Hilpert[1], A. Hoelzer[1], K. Linzmeier[1], C. Spenger [1] and P. Kroon[2]

[1] Fraunhofer Institute for Integrated Circuits IIS, 91058 Erlangen, Germany

[2] Agere Systems, Allentown, PA 18109, USA

ABSTRACT

Recently, a new approach in low bitrate coding of stereo and multi-channel audio has emerged: Spatial audio coding permits an efficient representation of multi-channel audio signals by transmitting a downmix signal along with some compact spatial side information describing the most salient properties of the multi-channel sound image. Besides its impressive efficiency allowing multi-channel sound at total bitrates of only 64 kbit/s and lower, the approach is also backward compatible to existing transmission systems and thus accommodates a smooth transition towards multi-channel audio in the consumer market. The paper gives an overview of the basic concepts and the options provided by spatial audio coding technology. It reports about some recent performance data, first commercial applications and related activities within the ISO/MPEG standardization group.

1. INTRODUCTION

Recently, a new approach in perceptual coding of multi-channel audio has emerged: Spatial audio coding technology extends traditional approaches for joint stereo coding of two or more channels in a way that provides several significant advantages, both in terms of compression efficiency and user features. Firstly, it allows the transmission of multi-channel audio signals at bitrates which so far have been used for the transmission of 2 channel stereo signals (or even monophonic signals). Secondly, by its underlying structure, the transmitted multi-channel audio signals is transmitted in a backward compatible way, i.e. spatial audio coding technology can be used to upgrade existing distribution infrastructures for stereo or mono audio content (radio channels, Internet streaming, music downloads, etc.) towards the delivery of multi-channel audio while retaining full compatibility with existing receivers. This paper describes the basic concepts behind the idea of spatial audio coding and outlines its relation to existing schemes for reproduction of multi-channel audio using transmission of non-multi-channel material. It provides some recent performance data of this type of technology and describes first commercial applications. Ongoing activities of the ISO/MPEG standardization group in this field will be reported.

2. PRINCIPLE & OVERVIEW

For many years, the idea of reproducing multi-channel audio based on transmitted stereo signals has been used to create a multi-channel like sound experience for the users when the delivery of discrete multi-channel audio was not available. This was achieved using several classes of schemes:

- Unguided creation of multi-channel sound: In order to provide a multi-channel experience with the pleasant sensation of envelopment, a number of techniques can be employed to convert an existing stereo signal into multi-channel which can be played back on home theater setups [1]. Many consumer receivers include such schemes which may involve decorrelation techniques and room/ambience modeling. While such schemes can enhance the user envelopment experience, they are by definition not able to recreate a desired multi-channel sound image as intended by a sound engineer since this information is not available to the scheme. In this sense, such schemes can be called "blind" upmixing.

- Matrixed surround techniques: For a long time, the delivery of multi-channel audio to consumers was only possible via analog stereo distribution channels, including Video Cassette Recorders (VCRs) and radio. This has been achieved by encoding multi-channel signals into stereo signals using phase shifting and adaptive sum/difference techniques and is now a part of practically all consumer home theater receivers (Dolby Surround/Prologic, Logic 7 …) [2]. While this type of technology provides indeed an approximation of the original multi-channel sound image, the encoding/decoding process is known to impose a number of limitations which result in a multi-channel sound quality which is clearly inferior to the quality of a full discrete delivery of multi-channel sound [3]. This forces sound engineers to take into account the limitations of matrixed surround schemes at the time of sound production in order to optimize the multi-channel sound quality delivered.

- Spatial audio coding techniques: In contrast to the matrixed surround approach, spatial audio coding techniques transmit somecompact spatial side information (e.g. 16 kbit/s for 5.1 content) in addition to the basic audio signal. This side information captures the most salient perceptual aspects of the multi-channel sound image, including level differences, time/phase differences and inter-channel correlation/coherence cues. As a consequence, such schemes do not have to rely anymore on the manipulation of signal phases for encoding spatial information. This makes it possible to even use a single (monophonic) audio channel as the basis for recreating multi-channel after spatial rendering.

The remainder of this paper will explain the concepts behind spatial audio coding, including mono-based and stereo-based rendering of multi-channel audio. This will be done by following the evolution of traditional techniques for joint stereo coding of multiple audio channels towards spatial audio coding techniques.

3. FROM JOINT STEREO CODING TO SPATIAL AUDIO CODING

3.1. Joint Stereo Coding

Joint-stereo coding denotes audio coding techniques which code two (or more) audio channels jointly in order to achieve a higher coding efficiency than would be possible by separate coding of the channels. Generally speaking, both inter-channel redundancy and inter-channel irrelevance is exploited for reducing the bitrate. In the following two most commonly used such techniques are described.

M/S stereo coding was introduced for low bitrate audio coding in [4]. A matrixing operation similar to the approach used in FM stereo transmission is used in the encoder with the appropriate dematrixing in the decoder. Rather than transmitting the left and right signal, the normalized sum and difference signals are used which are referred to as the middle (M) and the side (S) channel. The matrixing (i.e. sum/difference) operation is carried out on the spectral coefficients of the channel signals and can be thus performed in a frequency selective fashion. M/S stereo coding can be seen as a special case of a main axis transform of the input signal, rotating the input signals by a fixed angle of 45 degrees [5]. M/S coding is applied only when it is more efficient than coding the original signal channel pair. This is the case when the original signal pair is correlated and/or when the masked threshold for M/S coding is higher. (Due to the phenomenon of binaural masking level difference [6][7][8] the masked threshold depends on whether M/S coding is used or not).

Whithin the family of ISO/MPEG Audio coders, M/S stereo coding has been used extensively within the well-known MPEG-1/2 Layer 3 ("MP3") (full band on/off

switching) and within the MPEG-2/4 Advanced Audio Coder [9] in an enhanced fashion (individual switching for each scalefactor band). For use with multi-channel audio, M/S stereo coding is applied to channel pairs that are symmetric to the listener (front/back) axis.

Another commonly used joint stereo coding technique is "intensity stereo coding" (ISC) [5][10]. ISC exploits the fact that the perception of high frequency sound components mainly relies on the analysis of their energy-time envelopes [6]. Thus, it is possible for certain types of signals to transmit a single set of spectral values that is shared among several audio channels with only little loss in sound quality. The original energy-time envelopes of the coded channels are preserved approximately by means of scaling the transmitted signal to a desired target level which needs to be carried out individually for each frequency (scalefactor) band. Due to the underlying principle, use of ISC is commonly restricted to the high frequency range of the audio signal in order to avoid severe artifacts in the stereo image, especially for signals with a wide stereo image composed of decorrelated components, such as applause [7].

Within the family of ISO/MPEG Audio coders, ISC has been used both for all MPEG-1/2 coders as well as within the MPEG-2/4 Advanced Audio Coder [9]. For multi-channel audio coding, this coder can apply ISC in a generalized way by combining the spectral coefficients of two or more audio channels into a single set of spectral coefficients plus scaling information for each channel.

3.2. Parametric stereo

As a next step in the evolution of joint stereo perceptual audio coding, parametric stereo coding techniques have been proposed recently [11] [12] which further develop the basic idea of ISC coding to overcome many of its original limitations:

- Rather than using the coder's own filterbank, a dedicated (complex-valued, not critically-sampled) filterbank is used to re-synthesize the two channel stereo output from the transmitted mono channel. This avoids artifacts due to aliasing resulting from the spectral modification carried out for generating the output channels.

- Besides intensities, also phase differences and coherence cues between output channels are re-created. Coherence cues are a similarity measure between the channel pair.

Due to these improvements (in contrast to ISC), parametric stereo schemes can operate on the full audio bandwidth and thus can convert a monophonic signal coded by a base coder into a stereo signal. While development of such technology has originally been pursued in the context of the MPEG-4 parametric audio coder [13], the parametric stereo tool defined in this standard may also be applied in the context of the MPEG-4 HE AAC coder [14]. A detailed description of the parametric stereo tool is outside the scope of our discussion, for more information see e.g. [15].

3.3. Binaural Cue Coding (BCC)

Although predating parametric stereo in publication history, the *Binaural Cue Coding* (BCC) approach [16][17][18] can be considered a generalization of the parametric stereo idea, delivering multi-channel output (with an arbitrary number of channels) from a single audio channel plus some side information. Figure 1 illustrates this concept. Several input audio channels are combined into a single output ("sum") signal by a downmix process. In parallel, the most salient inter-channel cues describing the multi-channel sound image are extracted from the input channels and coded compactly as BCC side information. Both, sum signal and side information, are then transmitted to the receiver side, possibly using an appropriate low bitrate audio coding scheme for coding the sum signal. Finally, the BCC decoder generates a multi-channel output signal from the transmitted sum signal and the inter-channel cue information by re-synthesizing channel output signals which carry the relevant inter-channel cues, such as Inter-channel Time Difference (ICTD, Inter-channel Level Difference (ICLD) and Inter-channel Coherence (ICC). Figure 2 shows the general structure of a BCC synthesis scheme. The transmitted ("sum") signal is mapped to a spectral representation by a filterbank. For each output channel to be generated, individual time delays and level differences are imposed on the spectral coefficients, followed by a coherence synthesis process which re-introduces the most relevant aspects of coherence/ (de)correlation between the synthesized audio channels. Finally, all synthesized output channels are converted back into a time domain representation by inverse filterbanks. Since a detailed description of the BCC approach is beyond the scope of this paper, the reader is referred to [19] for a recent treatment of this technique.

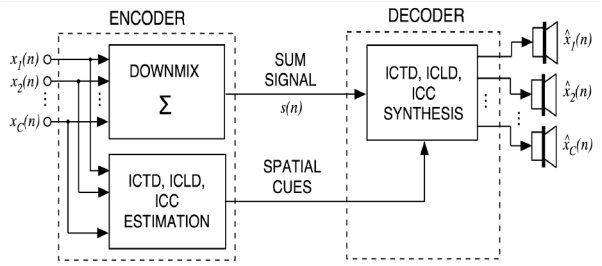

Figure 1: Principle of Binaural Cue Coding.

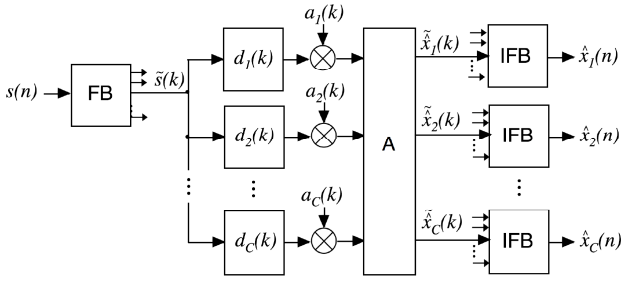

Figure 2: Binaural Cue Coding Synthesis (Principle).

Similar to parametric stereo, BCC exhibits a number of advantages over ISC by being able to recreate output signals with time differences and coherence cues (for recreating correct apparent source widths and ambience/listener envelopment). Consequently, BCC can be applied to the full audio frequency range without unacceptable signal distortions. Conversely, the traditional ISC processing can be interpreted as a BCC type processing which is limited to ILD synthesis only and is subject to imperfect reconstruction due to the use of critically subsampled coder filterbanks. An alternative type of BCC has also been used to enable bitrate-efficient transmission and flexible rendering of multiple audio sources which are represented by a single transmitted audio channel plus some cue side information [17][18] . The audio signal transmitted as part of the BCC scheme (see Figure 1) can be considered a backward compatible monophonic version of the multi-channel signal and thus be reproduced via conventional monophonic playback even without a BCC decoder. Thus, BCC effectively provides *mono compatible (mono-based) coding of multi-channel signals*.

3.4. C-to-E BCC and Spatial Audio Coding

The paradigm of BCC can be generalized. As opposed to transmitting a single audio channel, C-to-E BCC transmits E audio channels (usually E < C). A generic C-to-E BCC scheme is shown in Figure 3. As can be seen, it is identical to "C-to-1" BCC (Figure 1) except that more audio channels are transmitted from the encoder to the decoder. The E transmitted channels are generated by means of downmixing the C input channels. For example, 5.1 surround signals can be downmixed to two channels as suggested in [1].

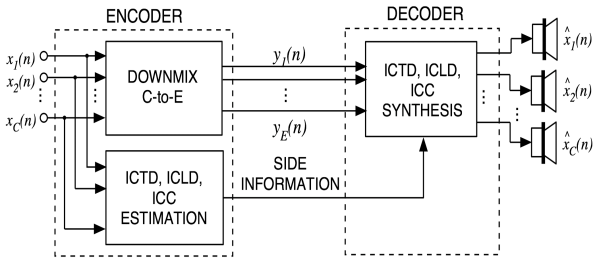

Figure 3: Generic scheme for C-to-E BCC.

Figure 4 shows the synthesis scheme at the decoder which generates the C output channels given the E transmitted channels. Note that this is very similar to "1-to-C" synthesis, except that prior to applying delays, scaling, and de-correlation (Processing Block A) an individual base channel is computed for each output channel as a linear combination of the transmitted channels ("C-to-1" BCC uses the same base channel for all output channels as indicated in Figure 2). The computation of the base channels is denoted upmixing and can be carried out in the time domain or spectral domain.

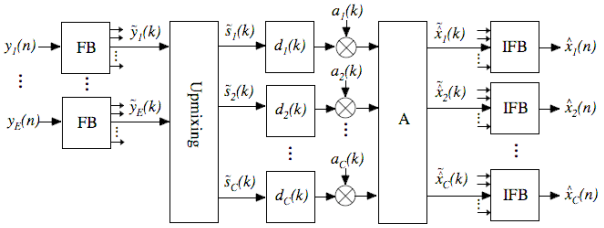

Figure 4: C-to-E BCC Synthesis.

C-to-E BCC is discussed in detail in [20]. An important special case of C-to-E BCC is *stereo compatible (stereo-based) coding of multi-channel signals* (i.e. C-to-2 BCC) where the two transmitted audio channels form a backward compatible stereo downmix for presentation via conventional stereo playback [3]. C-to-E BCC is, however, not only applicable to mono or stereo compatible coding of multi-channel surround audio, but also can be used for extending existing

surround formats towards more independent playback audio channels.

In view of the very broad and generic nature of these ideas, and consistent with the working title of a new standardization activity (see Section 6), we will refer to these concepts as the *Spatial Audio Coding* approach in the following.

4. RECENT PERFORMANCE RESULTS

The spatial audio coding approach is a very recent development within the field of multi-channel audio coding, and only a limited amount of performance data is currently available to characterize the potential (and possible limitations) of this concept. As can be expected from the underlying principle, the use of spatial cue side information should improve the subjective sound quality of spatial audio systems beyond what can be achieved by matrixed surround systems that work without such information. First test results were published recently for a stereo-based system based on 2-channel transmission and Enhanced Binaural Cue Coding ("MP3 Surround", see applications section) and indicated promising results [3]. Subsequently, results of more recent tests are presented which complement the previous data in two ways:

- How does the quality of a mono-based system compare to that of a stereo-based counterpart? From looking at the underlying principle, it can be expected that transmitting a compatible stereo downmix signal already conveys a significant part of the signal's original spatial characteristics and may thus lead to a better multi-channel output signal than can be achieved by a purely parametric reconstruction of the multi-channel sound from a single transmitted channel. On the other hand, a sophisticated mono-based scheme may be able to minimize this gap.

- Inclusion of and comparison to well-known matrixed surround technology into such listening tests plays an important role in the characterization of spatial audio coding systems, as this allows to relate the system's performance to a well-understood anchor signal and assess the actual improvement due to the use of side spatial side information. (The latter only applies to the case of stereo-based systems since there is no matrix-only counterpart for mono-based spatial audio coding systems).

The following sections briefly discuss issues of test methodology for critical multi-channel sound quality assessment and present some recent test results for stereo-based and mono-based representation and in this way illustrate the potential behind the spatial audio coding concept. These results are based on an evaluation of the EBCC (= Enhanced BCC) system, as it is part of the ongoing MPEG activities on this topic (see section on standardization) and represent a snapshot in time within the ongoing development process.

4.1. Test Methodology

To perform rigorous testing of multi-channel sound quality, it is critically important to use a suitable test methodology which also fulfills the following requirements:

- The method should allow a reliable and consistent ranking of different systems while remaining efficient in terms of listening effort per test subject.

- Comparing spatial sound image of several signals is particularly difficult, given the limited memory of human listeners for this type of sensation. Thus, a good listening test approach should take this into account and be sensitive to distortions and undesired alterations of the auditory spatial image.

Therefore, a listening test method was chosen in line with the ITU recommendation BS.1534 (MUSHRA) [21]. Several time-aligned audio signals were presented to the listener who performed on-the-fly switching between these signals using a keyboard and a screen. The signals included the original signal, which was labeled as "Reference", and several anonymized items, arranged in random order. Using a graphical user interface, the test subjects had to grade the basic audio quality of the anonymized items on a scale with five equally sized regions labeled "Excellent", "Good", "Fair", "Poor" and "Bad". To check the listener's reliability and enable relating the ratings to results of other tests, a hidden reference (original) and an anchor were included in addition to the coded/decoded items. The anchor consisted of a 3.5 kHz bandwidth reduced version of the reference, as is compulsory for the MUSHRA test methodology. There was no limit of the number of repetitions the test subjects could listen to before rating the item, and proceeding to the next test item. Furthermore, the listener was free to set start and stop markers for looping at arbitrary positions, as is suggested in the BS.1116-1 [22] test specification and

can be used for MUSHRA tests likewise. Due to its interactive nature, the test was taken by one subject at a time. As an additional anchor signal, Dolby Prologic 2 encoded/decoded multi-channel material was prepared and included as described in the annex.

Eight items were selected for the listening test, representing a mixture of music items of different styles (contemporary music, classical music, speech) and critical test material (applause signals, panned material, glockenspiel). Details on test material and listening test setup can be found in the annex of this paper.

4.2. Stereo-Based Spatial Audio Coding

Nine listeners participated in the test, seven of them being expert listeners with years of experience in audio coding. In order to capture the combined effects of low bitrate audio coding and spatial audio coding, as they would occur realistically in applications, the compatible stereo signal was encoded / decoded using AAC-LC at a bitrate of 128 kbit/s. An average spatial side information data rate of 19.1 kbit/s was consumed by the EBCC scheme for the given test set. The test results can be seen in Figure 5.

It can be observed that the performance of the EBCC scheme was better that the Prologic 2 anchor for all items in terms of mean performance, and was significantly better in a statistical sense (i.e. no overlap in the 95% confidence intervals) for 6 out of 8 items. This is a clear indication of how the use of side information improves sound quality beyond what is possible with a purely matrixed surround approach.

4.3. Mono-Based Spatial Audio Coding

Ten listeners participated in the test, eight of them being expert listeners with years of experience in audio coding. The compatible monophonic signal was encoded / decoded with an AAC-LC codec at 64 kbit/s. An average spatial side information data rate of 14.9 kbit/s was consumed by the EBCC scheme for the given test set. The test results are shown in Figure 6.

As expected, it can be seen that indeed reducing the number of transmitted channels from two to one results in a less perfect reconstruction of the multi-channel signals. While the overall performance of the EBCC scheme is still somewhat better in terms of grand mean, none of the items is significantly better in a statistical sense (i.e. no overlap in the 95% confidence intervals). Note that Prologic 2 was used without coding the transmitted stereo signals, whereas mono-based spatial audio coding was used with a 64kbit/s mono audio coder resulting in additional signal degradations. Thus, this result indicates that mono-based spatial audio coding can achieve sound quality at least as good as matrixed stereo surround schemes.

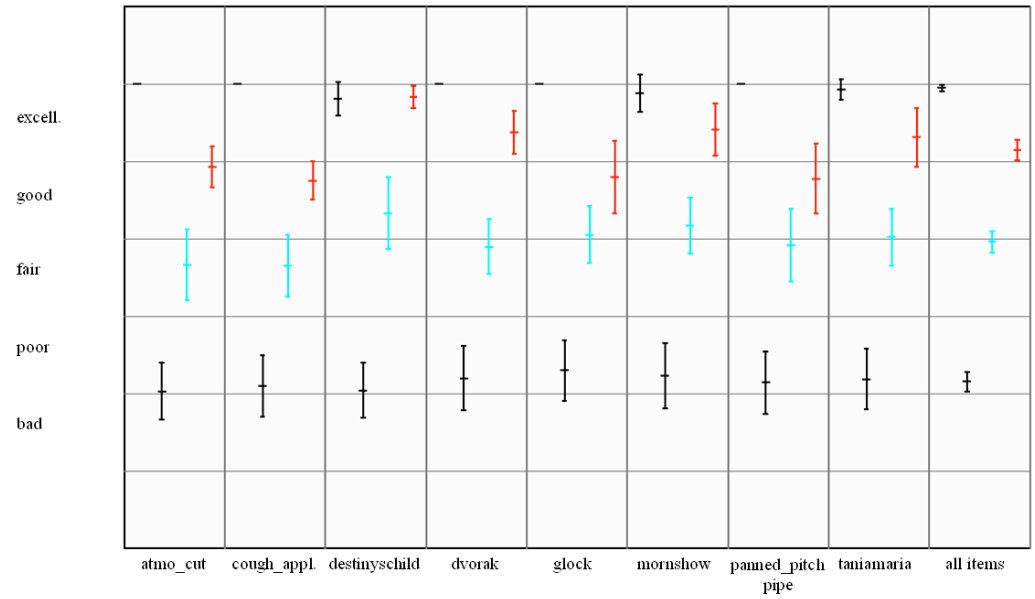

Figure 5: Performance of stereo-based system (mean grades and 95% confidence intervals).

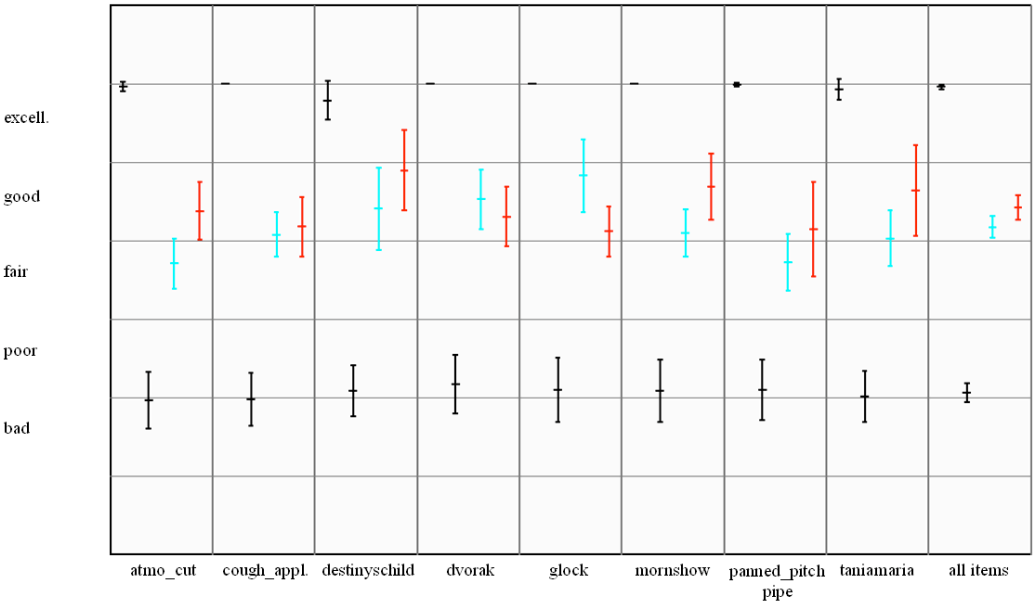

Figure 6: Performance of mono-based system (mean grades and 95% confidence intervals).

5. APPLICATIONS

Considering the general trend towards surround sound in consumer and professional audio, the following section illustrates the range of applications that are enabled through spatial audio coding, focusing on compatible multi-channel enhancements of existing services. After a general overview of possible application areas, two prominent examples of such applications are discussed in more depth, i.e. MP3 Surround and compatible digital broadcasting of multi-channel sound.

5.1. Application Areas

The current audio-visual distribution infrastructure is mainly tailored to delivering stereo rather than multi-channel audio, both in terms of available transmission bandwidth as well as the underlying technical system structure. Thus, in order to enable upgrading such distribution media to multi-channel audio, it is essential to deliver multi-channel audio at bitrates comparable to what is usually needed for the transmission of stereo which is one of the key features of the MP3 Surround/spatial audio coding approach.

- *Music download service*: Currently, a number of commercial music download services are available and working with considerable commercial success. Such services could be seamlessly extended to provide multi-channel enabled services while staying compatible for stereo users. On computers with 5.1 channel setups the compressed sound files are decoded in surround sound while on portable players the same files are played back as stereo music.

- *Streaming music service / Internet radio*: Many Internet radios are operating currently under severely constrained bandwidth conditions and, therefore, can offer only mono or stereo content. Spatial audio coding technology could extend this to a full multi-channel service within the permissible range of bitrates. Since efficiency is of paramount importance in this application, the compression aspect of these systems comes into play. As an example, representation of 5 presentation channels from two transmitted basis channels corresponds to a bitrate saving of (5-2)/5 = 60% compared to full multi-channel coding, neglecting the small amount of surround enhancement side information.

- *Digital Audio Broadcasting*: Due to available channel capacity, the majority of existing or planned systems for digital broadcasting of audio content cannot provide multi-channel sound to the users. Adding this feature could be a strong motivation for users to make the transition from their traditional FM receivers to the new digital systems. Again, compatibility with existing stereo reproduction setups is mandatory and inherent to the spatial audio coding concept.

- *Teleconferencing*: Although teleconferencing is becoming increasingly popular and important in today's global business world, most systems are still working with rather low bandwidth. A spatial audio coding approach can help expand the sound image to multi-channel sound and, thus, allow a better subjective separation and resolution of the audio contributions of each speaker in the teleconference.

- *Audio for Games*: Many personal computers have become "personal gaming engines" and are equipped with a 5.1 computer speaker setup. Synthesizing 5.1 sound from a backward compatible stereo sound basis allows for an efficient storage of multi-channel background music.

For multi-channel applications requiring the lowest possible bitrate, a spatial audio coding approach based on a mono channel can be used. For 5 channels, a bitrate saving of about 80% compared to a discrete multi-channel transmission can be achieved.

5.2. MP3 Surround

A first commercial application of the Spatial Audio Coding idea has recently been described under the name *MP3 Surround* and is based on the well-known MPEG-1/2 Layer 3 algorithm as an audio coder for transmission [3]. Figures 7 and 8 illustrate the general structure of MP3 Surround encoding / decoding for the case of a 3/2 multi-channel signal (L, R, C, Ls, Rs). As a first step, a two-channel compatible stereo downmix (Lc, Rc) is generated from the multi-

channel material by a downmixing processor (or as a separately produced artistic downmix). The resulting stereo signal is encoded by a conventional MP3 encoder in a fully standards compliant way. At the same time, a set of spatial parameters (ICLD, ICTD, ICC) is extracted from the multi-channel signal, encoded and embedded as surround enhancement data into the ancillary data field of the MP3 bitstream. On the decoder side, the MP3 Surround bitstream is decoded into a compatible stereo downmix signal that is ready for presentation over a conventional 2-channel reproduction setup (speakers or headphones). Since this step is based on a compliant MPEG-1 Audio bitstream, any existing MP3 decoding device can perform this step and thus produce stereo output. MP3 Surround enabled decoders will detect the presence of the embedded surround enhancement information and, if available, expand the compatible stereo signal into a full multi-channel audio signal using a BCC-type decoder.

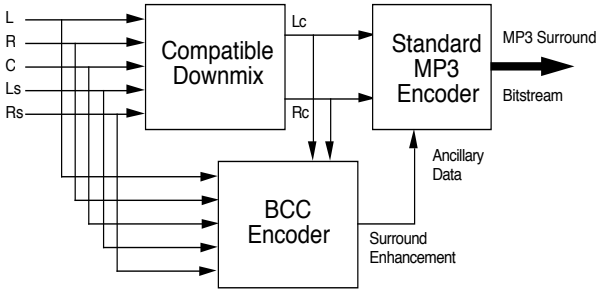

Figure 7: Principle of MP3 Surround Encoding [3].

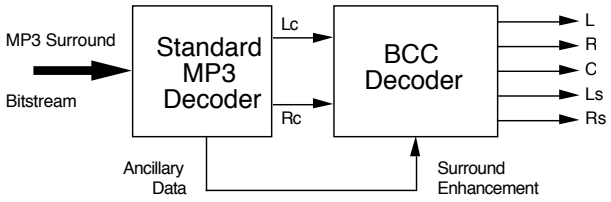

Figure 8: Principle of MP3 Surround Decoding [3].

5.3. Compatible 5.1 Broadcasting

Over the recent decade, several digital audio broadcasting (DAB) formats have been introduced, both terrestrial and satellite-based [23]. Two main digital terrestrial services intended to replace or enhance the current analog broadcasting services are the Eureka-147 [24] system deployed mainly in Europe and Asia, and HD Radio as proposed by Ibiquity in the US [25]. Other digital broadcasting efforts are being developed by Digital Radio Mondiale (DRM) to replace the current mid- and short-wave analog transmissions [26]. In the US, two satellite broadcasters have been in commercial operation since 2002: Sirius [27] and XM [28], with a total of about 2.5 million subscribers.

The success of these new digital radio services will depend heavily on the perceived value and associated costs for the end user. While the value proposition for the consumer is quite clear for the satellite services (they provide a large selection of program material, most of the programs are commercial free, and the coverage extends over a very large geographical area), the same is not necessarily true for the terrestrial digital broadcasting services. In essence, these digital systems replace the underlying analog transmissions systems without changing the program format and content. Depending on the system, there might be some additional advantages such as the Single Frequency Network (SFN) concept in Eureka-147, which extends coverage, and improved audio quality specifically for the systems replacing AM and short-wave transmission technologies. However, the majority of listeners are currently listening to FM based analog systems with a reasonable coverage and audio quality, and to them the benefit of the digital systems appears limited. Even for those listeners that perceive value from the digital format, there is an additional barrier in terms of additional cost of the receiver. Therefore other arguments and features may need to be provided to make digital radio a compelling story.

Spatial Audio Coding technology, as described in this paper, could be a key factor in boosting the attraction of digital radio systems since it provides additional functionality not obtainable in other ways. This is specifically true for the surround sound capabilities enhancing current stereo presentation towards multi-channel. The capability of delivering surround sound appears particularly appealing since this format has been steadily gaining more and more acceptance in many households over the recent years. While multi-channel sound is primarily reproduced via “home theatre” setups, typical households will increasingly use the same setups for their regular radio playback and thus have the “playback infrastructure” for multi-channel reproduction.

A second strong potential appeal of multi-channel sound is playback in the car – considering the facts that most radio listening occurs in cars, and the automotive environment is well suited for enjoying multi-channel music in terms of placement of speakers and stable listener position relative to the loudspeakers. Audio is the time-tested accompaniment to driving and surround is a natural next step. Multi-channel playback capability in cars is about to emerge quickly. Most cars already have 4 (or more) loudspeakers, even though they are used to reproduce two-channel FM broadcasts. The newer generation of cars is expected to have 5.1 loudspeakers setups in the near future, luxury cars come standard equipped with 5.1 speaker setup and a DVD-Audio or SACD player today.

The reasoning above seems to suggest that surround sound for terrestrial digital radio broadcasts can be considered a killer application by providing, in addition to the existing services, a new benefit, which is clearly noticeable. Due to this increased value perception, it is possible to justify somewhat increased costs for the digital receiving equipment. From a broadcaster perspective, the additional burden of transmitting a spatial audio based format is insignificant, since the low bitrate spatial side information can be easily transmitted using existing data fields in the bitstream. Together with the inherent backward compatibility of the spatial audio coding approach, this provides the prerequisites for a seamless transition from stereo (or mono) to multi-channel broadcasting accommodating all types of users, existing equipment and services.

First trials in exploiting the potential of spatial audio coding in the context of digital audio broadcasting are on their way REF[29].

6. STANDARDIZATION

In the area of international standardization, the ISO/MPEG Audio group has become aware of the recent advances in BCC–based spatial audio coding technology, and a number of demonstrations were observed illustrating the benefits of combining such technology with different MPEG source coding algorithms:

- MPEG-1 Layer 3 + EBCC: 5.0 multi-channel sound at 192 kbit/s (115th AES Convention, NY, 10/2003, and 66th MPEG meeting, Brisbane, 10/2003)
- MPEG-4 AAC + EBCC: 5.1 multi-channel sound at 140 kbit/s (67th MPEG meeting, Hawaii, 12/2003)
- MPEG-4 High Efficiency AAC + EBCC: 5.1. multi-channel sound at 48 kbit/s (NAB Conference and Exhibition 2004, Las Vegas, 3/2004, and 68th MPEG Meeting, Munich, 3/2004)

In view of these recent advances and their market potential [30] ISO/MPEG Audio started a new work item on *Spatial Audio Coding*. This process aims at complementing the existing MPEG-4 AAC-based general audio coding schemes (but also other formats such as PCM) with a tool for efficient and compatible representation of multi-channel audio. It addresses both technology that expands stereo signals into multi-channel sound (called "2-to-n" scheme) and the more traditional mono variant (called "1-to-n" scheme). The key requirements can be paraphrased as follows [31]:

- Best possible approximation of original perceived multi-channel sound image
- Minimal bitrate overhead compared to conventional transmission of 1 or 2 audio channels
- Backward compatibility of transmitted audio signal with existing mono or stereo reproduction systems, i.e. the transmitted audio channels shall represent a compatible (mono or stereo) audio signal representing all parts of the multi-channel sound image
- Independence from audio codec (among other transmission schemes the technology is expected to support MPEG-4 AAC and HE-AAC profile coders)
- Single unified architecture for both "1-to-n" and "2-to-n" processing

As of the time of writing of this paper, the MPEG group has issued a "Call for Proposals" (CfP) at its 68th meeting in March 2004 [31]. Four submission in response to this call were received at the July meeting and undergo competitive evaluation. The selection of the first Reference Model (RM) is scheduled for

October 2004 which will be the baseline for the subsequent development process.

7. CONCLUSIONS

Spatial audio coding technology is the most recent addition in the family of technologies which deliver multi-channel audio to consumers via non-multi-channel audio channels. In contrast to a fully discrete transmission, this enables the delivery of multi-channel sound at bitrates of 64 kbit/s and lower. This efficiency in representing multi-channel sound and its inherent backward compatibility to mono or stereo audio transmission make spatial audio coding a promising technology for the migration of today's distribution infrastructures towards multi-channel audio. A broad range of applications may benefit from the spatial audio coding approach. Standardization efforts in this field have been started.

8. ANNEX

The subjective listening test was conducted in an acoustically isolated listening lab that is designed to permit high-quality listening tests conforming to the BS.1116 [22] test methodology for high-quality audio listening tests. The room dimension is 5.5m x 7.1m x 2.4m. All test signals were presented on 5 Geithain RL 901 active studio loudspeakers and a Geithain TT 920 active subwoofer driven by Bryston pre-amplifiers. Playback was controlled from a 2 GHz Linux computer with an RME Hammerfall digital sound output interface connected to Lake People DAC F20 D/A converters.

The Dolby ProLogic II signals were generated with the *SurCode for Dolby ProLogic II Version 2.0.3* software by Minnetonka Audio Software using the default settings. In order to guarantee perceptually equal loudness of the Dolby ProLogic II signals, a loudness alignment procedure was necessary. The amplification factor for each ProLogic II decoded signal was determined by adjusting its amplification in steps of 1 dB until there was no perceptual difference in loudness compared to the original signal when switching between both signals. The procedure was repeated for several listeners. This resulted in an equal amplification of 3 dB for all ProLogic II decoded test signals. The signals had sufficient headroom to enable this gain change without introducing clipping.

The following test material was used:

Item	Format	Source	Genre
atmo_cut	5.0	Multi-Channel Demo Material (Denon Electronic GmbH)	applause
cough_ applause	5.0	AT&T	applause
destinyschild	5.1	SACD Destiny´s Child: Survivor	contemporary music
dvorak	5.0	SACD Iván Fischer: Dvorak Slavonic Dances	classical music
glock	5.0	MPEG Multi-Channel Audio Material	instruments
mornshow	5.0	Multi-Channel Demo Material (Telos Systems)	speech
panned_ pitchpipe	5.0	Created from MPEG Multi-Channel Audio Material	artificial
taniamaria	5.1	SACD Tania Maria: Come With Me	contemporary music

9. REFERENCES

[1] Rec. ITU-R BS.775 (1993), Multi-Channel Stereophonic Sound System with or without Accompanying Picture, ITU. http://www.itu.org.

[2] Rumsey, F. (2001), Spatial Audio, Focal Press, Music Technology Series.

[3] J. Herre, C. Faller, C. Ertel, J. Hilpert, A Hoelzer, C. Spenger: "MP3 Surround: Efficient and Compatible Coding of Multi-Channel Audio", 116th AES Convention, Berlin 2004, Preprint 6049.

[4] J.D. Johnston: "Perceptual Coding of Wideband Stereo Signals", Proc. of the ICASSP 1990.

[5] R.G.v.d. Waal, R.N.J. Veldhuis: "Subband Coding of Stereophonic Digital Audio Signals", IEEE ICASSP 1991, pp. 3601–3604.

[6] J. Blauert, "Spatial Hearing: The Psychophysics of Human Sound Localization", revised edition, MIT Press, 1997.

[7] "Perceptual Audio Coders: What to Listen For", Demonstration CD-ROM on Audio Coding Artifacts, AES Publications, 2002.

[8] J. Herre, K. Brandenburg, E. Eberlein: "Combined Stereo Coding", 93rd AES Convention, San Francisco 1992, Preprint 3369.

[9] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, Oikawa, "ISO/IEC MPEG-2 Advanced Audio Coding", Journal of the AES, Vol. 45, No. 10, October 1997, pp. 789-814.

[10] J. Herre, K. Brandenburg, D. Lederer: "Intensity Stereo Coding", 96th AES Convention, Amsterdam 1994, Preprint 3799.

[11] E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart, "Advances in parametric coding for high-quality audio," 114th AES Convention, Amsterdam 2003, Preprint 5852.

[12] E. Schuijers, J. Breebaart, H. Purnhagen, J. Engdegård: "Low-Complexity Parametric Stereo Coding", 116th AES Convention, Berlin 2004, Preprint 6073.

[13] ISO/IEC JTC1/SC29/WG11 (MPEG), International Standard ISO/IEC 14496-3:2001/FDAM2, Parametric Coding, 2004.

[14] ISO/IEC JTC1/SC29/WG11 (MPEG), International Standard ISO/IEC 14496-3:2001/AMD1, Bandwidth Extension, 2003.

[15] H. Purnhagen: "Low Complexity Parametric Stereo Coding in MPEG-4", 7th International Conference on Audio Effects (DAFX-04), Naples, Italy, October 2004.

[16] C. Faller, F. Baumgarte: "Efficient Representation of Spatial Audio Using Perceptual Parametrization", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York 2001.

[17] C. Faller and F. Baumgarte, "Binaural Cue Coding: A novel and efficient representation of spatial audio," Proc. ICASSP 2002, Orlando, Florida, May 2002.

[18] C. Faller and F. Baumgarte, "Binaural Cue Coding - Part II: Schemes and applications," IEEE Trans. on Speech and Audio Proc., vol. 11, no. 6, Nov. 2003.

[19] C. Faller: "Parametric Coding of Spatial Audio", Swiss Federal Institute of Technology Lausanne (EPFL), Ph. D. Thesis No. 3062, 2004.

[20] C. Faller: "Coding of Spatial Audio Compatible with Different Playback Formats", 117th AES Convention, San Francisco 2004.

[21] ITU-R Recommendation BS.1534-1, "Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)", International Telecommunications Union, Geneva, Switzerland, 2001.

[22] ITU-R Recommendation BS.1116-1 "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems", International Telecommunications Union, Geneva Swizerland, 1994-1997.

[23] Technical Advances in Digital Audio Radio Broadcasting – C. Faller, B-H. Juang, P. Kroon, H-L. Lou, S.A. Ramprashad, C-E. Sundberg Proc. IEEE, August 2000, pp. 1303-1333.

[24] EUREKA147 - Digital Audio Broadcasting System. [Online]. Available: www.eurekadab.org.

[25] Ibiquity [Online] Available: www.ibiquity.com.

[26] Digital Radio Mondiale. [Online]. Available: www.drm.org.

[27] Sirius Satellite Radio [Online] Available: www.sirius.com.

[28] XM Satellite Radio [Online] Available: www.xmsr.com.

[29] "DAB 5.1 Surround via Spatial Audio Coding", Presentation at "Medientage 2004", Munich, Germany, October 2004.

[30] ISO/IEC JTC1/SC29/WG11 (MPEG), Document M10378, "Spatial Audio Coding: Market Context and Requirements", Hawaii 2003.

[31] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N6455, "Call for Proposals on Spatial Audio Coding", Munich 2004.